

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:09:13 ; Search time 299.73 Seconds

(without alignments)
16.815 Million cell updates/sec

Title: US-09-331-631a-3_COPY_74-116

Perfect score: 250

Sequence: 1 NQDDPQTDCQQCQRRCRQOE..... RQOQYCORRCKEICEEEYY 43

Scoring table: BLOSUM62

Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rabbit:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	250	100.0	666	10	Q9SP14	09SP14 macadamia i
2	242	96.8	666	10	Q9SP15	09SP15 macadamia i
3	235	94.0	625	10	Q9SP13	09SP13 macadamia i
4	115	46.0	593	10	Q9SEW4	09sew4 juglans reg
5	112	44.8	525	10	Q94338	043384 theobroma c
6	92	36.8	411	5	P91419	P91419 caenorhabdi
7	78	31.2	572	5	Q19594	019594 caenorhabdi
8	77.5	31.0	425	5	Q17400	017400 caenorhabdi
9	77.5	31.0	600	5	Q17401	017401 caenorhabdi
10	77	30.8	810	10	Q92W13	092w13 cucurbita m
11	73	29.4	5	Q19919	019919 caenorhabdi	
12	73	29.2	111	5	Q19054	019054 caenorhabdi
13	72	28.8	330	5	Q18118	018118 caenorhabdi
14	72	28.8	339	5	Q9n4r1	09n4r1 caenorhabdi
15	71	28.4	388	5	Q16500	016500 caenorhabdi
16	71	28.4	388	5	Q16501	016501 caenorhabdi
17	71	28.4	438	5	Q16502	016502 caenorhabdi
18	71	28.4	445	5	Q16511	016511 caenorhabdi
19	70.5	419	5	Q9na38	Q9na38 caenorhabdi	

ALIGMENTS

RESULT 1
Q9SP14 PRELIMINARY; PRT; 666 AA.
ID 09SP14; AC 09SP14; DT 01-MAY-2000 (TRIMBLErel. 13, Created)
DT 01-MAY-2000 (TRIMBLErel. 13, Last sequence update)
DT 01-OCT-2000 (TRIMBLErel. 15, Last annotation update)
DE VICILIN PRECURSOR.
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OC NCBI_TaxID=60983;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUT KERNEL;
RA Marcus J.P., Goultier K.C., Green J.I., Manners J.M.;
RT "A family of antimicrobial Peptides is Produced by processing of a 7S
RT globulin protein in Macadamia integrifolia.";
RL Plant. J. 0:0-0(1999).
DR EMBL; AF161884; AAD84245.1; -.
DR HSSP; P02053; PPHL.
DR INTERPRO; IPR00113; -.
DR PFAM; PF00546; Seedstore_7S; 1.
SEQUENCE 666 AA; 78243 MW; 0EC422F8710F8ATB CRC64;

Query Match Best Local Similarity 100.0%; score 250; DB 10; Length 666; Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NQDDPQTDCQQCQRRCRQEQSGPROQYCORRCKEICEEEYY 43
Db 74 NQDDPQTDCQQCQRRCRQEQSGPROQYCORRCKEICEEEYY 116
DT 01-MAY-2000 (TRIMBLErel. 13, Last sequence update)

Db	78	EEELQRYQOCQGRQEQQQCQRKWEQYKOE	117	RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RESULT	6			CC -1- SIMILARITY: BELONGS TO FAMILY UPF.
ID	P91419	PRELIMINARY;	PRT;	411 AA.
AC	P91419;			EMBL; U5197; AC48159.1; -.
DT	01-MAY-1997	(TREMBLrel. 03, Created)		DR WORMPEP; F19G12.7; CS07090.
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)		
DE	CODED FOR BY C. ELEGANS	CDNA YK115A6.5.		
GN	T01D1.6.			KW HYPOTHETICAL protein; Signal.
OS	Caenorhabditis elegans.			FT SIGNAL 1 21 POTENTIAL.
OC	Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;			FT CHAIN 22 572 MW; BDA52FDB850AAB9 CRC64;
OX	Rhabditidae; Peioderinae; Caenorhabditis.			SEQUENCE 572 AA; 62384 MW;
NCBI_TAXID=6239;				SQ
RN	{1}			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RX	MEDLINE#9450718; PubMed=7906398;			
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,			
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,			
RA	Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,			
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,			
RA	Jones M., Kershaw J., Kirstein J., Laister N., Latrice P.,			
RA	Lighthill J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,			
RA	Parsons J., Percy C., Riffkin L., Roopra A., Saunders D., Showman R.,			
RA	Simlton N., Smith A., Sonnauer E., Staden R., Sulston J.,			
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Watson R.,			
RA	Watson A., Weinstock L., Wilkinson-Sproat J., Wohlgemuth P.,			
RT	*2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans ";			
RL	Nature 368:32-38(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RA	Bradshaw H., Wohlgemuth P.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RA	Waterson R.;			
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.			
SQ	SEQUENCE 411 AA; 44675 MW; 599DAC9DFAB1382 CRC64;			
RESULT	7			RESULT 8
ID	Q19594	PRELIMINARY;	PRT;	425 AA.
AC	Q19594;			ID 017400 PRELIMINARY; PRT; 425 AA.
DT	01-JUN-1998	(TREMBLrel. 06, Created)		AC 017400;
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)		DT 01-JAN-1999 (TREMBLrel. 09, Created)
DE	HYPOTHETICAL	62.4 KDA PROTEIN F19G12.7 IN CHROMOSOME X PRECURSOR.		DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
GN	F19G12.7.			DE HYPOHETICAL 67.7 KDA PROTEIN AC3.4 IN CHROMOSOME V.
OS	Caenorhabditis elegans.			AC3.4.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			OS Caenorhabditis elegans.
OC	Rhabditidae; Peioderinae; Caenorhabditis.			OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OX	NCBI_TAXID=6239;			OC Rhabditidae; Peioderinae; Caenorhabditis.
RN	[1]			OX NCBI_TAXID=6239;
RP	SEQUENCE FROM N.A.			RN [1]
RC	STRAIN=BRISTOL N2;			RP SEQUENCE FROM N.A.
RA	Nhan M., Le T.T.;			RC STRAIN=BRISTOL N2;
RA				RA Nhan M., Le T.T.;
RESULT	7			RESULT 9
ID	Q19594	PRELIMINARY;	PRT;	600 AA.
AC	Q19594;			ID Q17401 PRELIMINARY; PRT; 600 AA.
DT	01-JUN-1998	(TREMBLrel. 06, Created)		AC 017401;
DT	01-JUN-1998	(TREMBLrel. 09, Last sequence update)		DT 01-JAN-1999 (TREMBLrel. 09, Created)
DE	HYPOTHETICAL	67.7 KDA PROTEIN AC3.4 IN CHROMOSOME V.		DT 01-JAN-2000 (TREMBLrel. 14, Last sequence update)
GN	F19G12.7.			DE HYPOHETICAL 67.7 KDA PROTEIN AC3.4 IN CHROMOSOME V.
OS	Caenorhabditis elegans.			OS Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Peioderinae; Caenorhabditis.			OC Rhabditidae; Peioderinae; Caenorhabditis.
OX	NCBI_TAXID=6239;			OX NCBI_TAXID=6239;
RN	[1]			RN [1]
RP	SEQUENCE FROM N.A.			RP SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;			RC STRAIN=BRISTOL N2;
RA	Nhan M., Le T.T.;			RA Nhan M., Le T.T.;

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO FAMILY UPF.
 DR EMBL: Z71177; CAA9468.1; -.
 DR WORMPP: AC3.4; CE05134.
 DR INTERPRO: IPR001504; -.
 DR PFAM: PF01529; zf-DHIC; 1.
 DR PRODOM: P000301; -! 1.
 FT Hypothetical protein; Transmembrane.
 FT TRANSMEM 309 329 POTENTIAL.
 FT TRANSMEM 447 467 POTENTIAL.
 FT TRANSMEM 490 510 POTENTIAL.
 FT DOMAIN 6 287 GLN-RICH.
 SQ SEQUENCE 600 AA; 67740 MW; 504A5CE1BA72091B CRC64;

RESULT 10
 Q9ZWI3 PRELIMINARY; PRT; 810 AA.
 ID Q9ZWI3; AC Q9ZWI3; DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE PV100
 OS Cucurbita maxima (Pumpkin) (Winter squash).
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Cucurbitales; Cucurbitaceae; Cucurbita.
 OC NCBI_TaxID=3661;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KUROKAWA_AMAKURI_NANKIN; TISSUE=COTYLEDON;
 RX MEDLINE=99107919; PUBMED=9891029;
 RA Yamada K., Shimada T., Kondo M., Nishimura M., Hara-Nishimura I.;
 RT "Multiple functional proteins are produced by cleaving Asn-Gln bonds
 of a single precursor by vacuolar processing enzyme.",
 RL J. Biol. Chem. 274:2563-2570(1999).
 DR EMBL: AB019195; BAA24056.1; -.
 DR HSSP: P02653; 2PHL.
 DR INTERPRO: IPR001113; -.
 DR PFAM: PF00546; SeedsToRe_S; 1.
 DR PRODOM: P0081059; -! 1.
 DR SEQUENCE 810 AA; 97314 MW; A829A3F7542266AB CRC64;

Query Match 30.8%; Score 77; DB 10; Length 810;
 Best Local Similarity 34.2%; Pred. No. 0.076; Mismatches 13; Conservatve 10; Indels 15; Gaps 0;
 Matches 13; Conservation 10; Mismatches 15; Indels 0; Gaps 0;

QY 5 P0TDCQOCQRRCRQQESGPRQQYQQRKEICEBEE 42
 ID :
 DB 74 PRAEVYEVCRURQVNERGYEQKKEQVEERLBBEQ 111

RESULT 11
 Q19919 PRELIMINARY; PRT; 242 AA.
 ID Q19919; AC Q19919; DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DE HYPOTHETICAL_26.2 KDA PROTEIN F31A3.1 IN CHROMOSOME X.
 GN F31A3.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peliaderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PUBMED=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kerswell J., Kursten J., Laister N., Latrellie P.,
 Lightning J., Lloyd C., McMurray K., Mortimore B., O'callaghan M.,
 Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Shownkeen R.,
 Smalton N., Smith A., Sonnhammer E., Straden R., Sulston J.,
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sprat J., Wohldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.",
 RL Nature 368:32-38(1994).
 DR EMBL: Z66496; CAA91281.1; -.
 DR SEQUENCE 111 AA; 11681 MW; AFR595F0DA154F33 CRC64;

Query Match 29.2%; Score 73; DB 5; Length 111;
 Best Local Similarity 41.9%; Pred. No. 0.047; Mismatches 13; Conservatve 3; Indels 15; Gaps 0;
 Matches 13; Conservation 3; Mismatches 15; Indels 0; Gaps 0;

QY 3 DDPQTDCQOCQRRCRQQESGPRQQYQQRKEICEBEE 33
 ID :
 DB 61 DDPNTDCTQYFLCSNAKYTPPLQOFCAKC 91

RESULT 13
 Q18118 PRELIMINARY; PRT; 330 AA.

AC 018118;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
 DE T23FL6. PROTEIN.
 GN OS
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderiniae; Caenorhabditis;
 OX NCBI_TAXID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilkinson J.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Showkeen R.,
 Shaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sprott J., Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans." [2]
 RA Nature 368:32-38(1994).
 DR EMBL; 281129; CAB03405:1; -.
 SQ SEQUENCE 330 AA; 36605 MW; F043B1A90D3A8FE9 CRC64;

Query Match 28.8%; Score 72; DB 5; Length 330;
 Best Local Similarity 29.5%; Pred. No. 0.15; Mismatches 9; Indels 12; Gaps 2;
 Matches 13; Conservative 10; MisMatches 9; Indels 12; Gaps 2;

QY 7 TPDQQCQCRRCRQQ-----ESGPQQQ--YCYCRCKEICE 38
 DB 230 TNCQQCQNSCQNSNTQTTIYQASQTSQVPCQPCQQCQCQ 273

RESULT 14

Q9N4R1 Q9N4R1 PRELIMINARY; PRT; 339 AA.
 ID Q9N4R1;
 AC 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOTHETICAL PROTEIN Y5H2A.C.
 GN Y5H2A.C.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderiniae; Caenorhabditis.
 OX NCBI_TAXID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX SEQUENCE FROM N.A.
 RA STRAIN=BRISTOL M2;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL M2;
 RA Waterston R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF016451; AAB6001:1; -.
 SQ SEQUENCE 388 AA; 42139 MW; 2E20655B0B9AE492 CRC64;

Query Match 28.4%; Score 71; DB 5; Length 388;
 Best Local Similarity 32.5%; Pred. No. 0.23; Mismatches 11; Indels 6; Gaps 2;
 Matches 13; Conservative 10; MisMatches 11; Indels 6; Gaps 2;

QY 6 OTDCQ-OCQRC-----RQOBGPROQYCORCKEICE 39
 DB 194 QAQCCPQCCQCCQSSCQVQQQDNOCEPAQVQOCSDICQ 233

None;
 "Genome sequence of the nematode *C. elegans*: a platform for investigating biology." The *C. elegans* Sequencing Consortium. ";
 Science 282:2012-2018(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=BRISTOL M2;
 RC Waterston R.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC006809; AAH59626:1; -.
 SQ SEQUENCE 339 AA; 36218 MW; 4023342323C432D CRC64;

Query Match 28.8%; Score 72; DB 5; Length 339;
 Best Local Similarity 32.7%; Pred. No. 0.15; Mismatches 8; Indels 18; Gaps 4;
 Matches 17; Conservative 9; MisMatches 8; Indels 18; Gaps 4;

QY 5 P0TDCQOCQRCR-----QOBGSPR-----QOYCORCKE--ICEFE 40
 DB 172 PQ--CQQCQNTQQFAPVCQQCAPCTISSAPQCQQCQT"QCQFAPVCQQ 221

RESULT 15

Q9N500 Q9N500 PRELIMINARY; PRT; 388 AA.
 ID Q9N500;
 AC 016500;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE C03A7.4 PROTEIN.
 GN CAenorhabditis elegans.
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderiniae; Caenorhabditis.
 OX NCBI_TAXID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL M2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Showkeen R.,
 Shaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sprott J., Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans." [2]
 RA Nature 368:32-38(1994).
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL M2;
 RA Greco T., Bradshaw H., Elliott G.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL M2;
 RA Waterston R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF016451; AAB6001:1; -.
 SQ SEQUENCE 388 AA; 42139 MW; 2E20655B0B9AE492 CRC64;

Search completed: March 1, 2001, 16:09:14
 Job time: 1558 sec

